POINT CROSS Life Sciences

Leveraging LLMs for Traceable Metadata Specifications and Executable Search Expressions in TFL & ADaM Generation

About PointCross

Origin (1999)
Founded 24 years ago, building big-data solutions. Headquarters: Foster city, CA

Focus (2009)
Exclusive focus on the BioPharma industry after acquisition of PharmQuest

Expansion (2015)
Nonclinical, FDA CDER, CTP.
eDV: Compliance & Quality for SEND, SDTM & ADaM

Clinical Footprint (2016)
Clinical Trail Data: Standardization of Biomarker & Translational research

CDISC SEND (2020)
One of the largest CDISC SEND provider

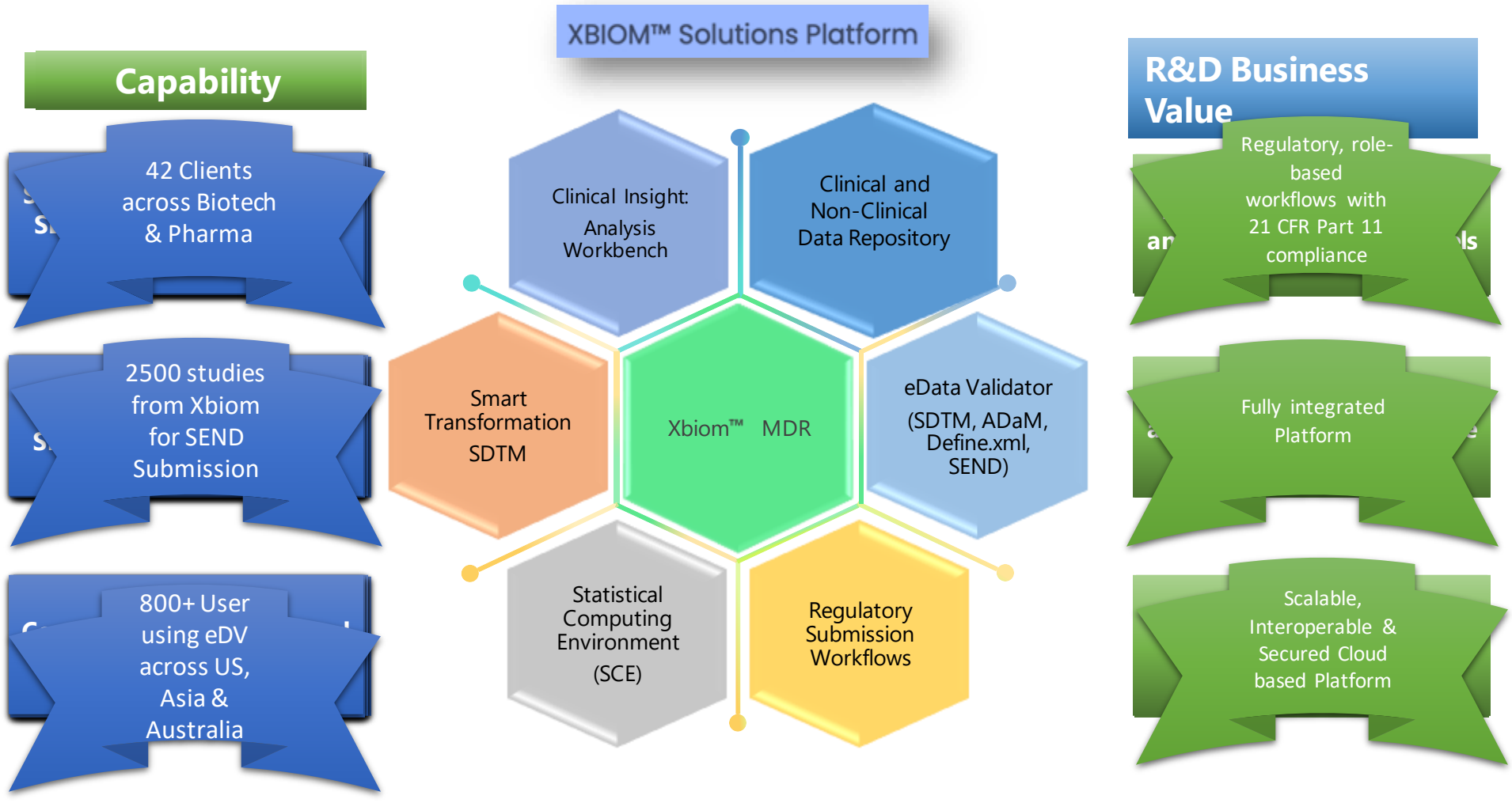
Xbiom™ : Platform (2023)
- Xbiom™ MDR
- Xbiom™ Smart Transform
- Xbiom™ Clinical Insight
- Xbiom™ eDV
- Xbiom™ SCE

Platinum member of CDISC and contributor to PhUSE

Serving 60+ Biotech & Pharma clients

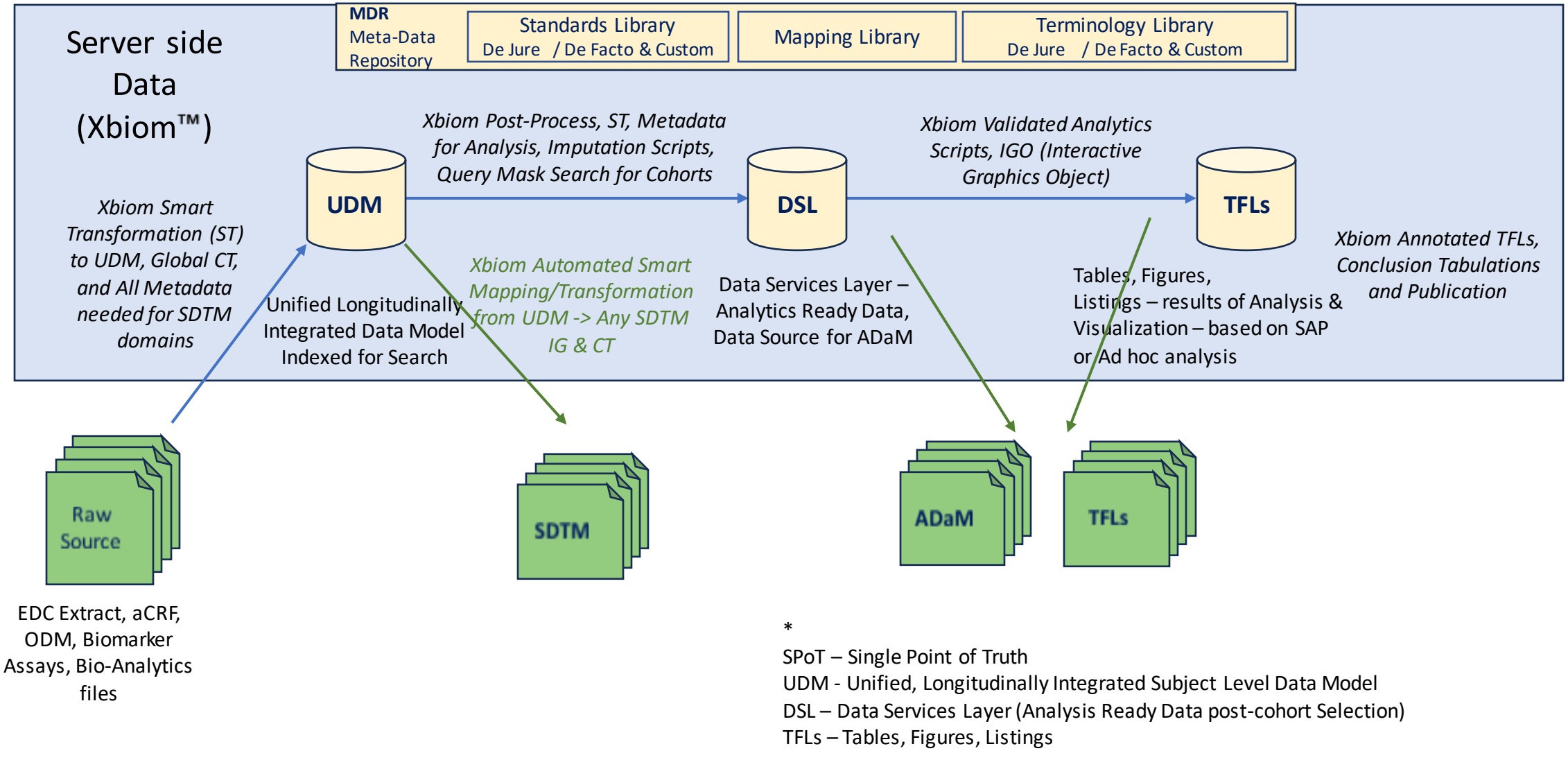
- ✓ Global Team: US, France & India
- ✓ 10000+ no. of users downloaded eDV across US, Asia, Europe & Australia
- ✓ Standardized 7,500+ studies for repository, analysis over 14 years

Xbiom™ Solutions Platform



Xbiom MDR and UDM

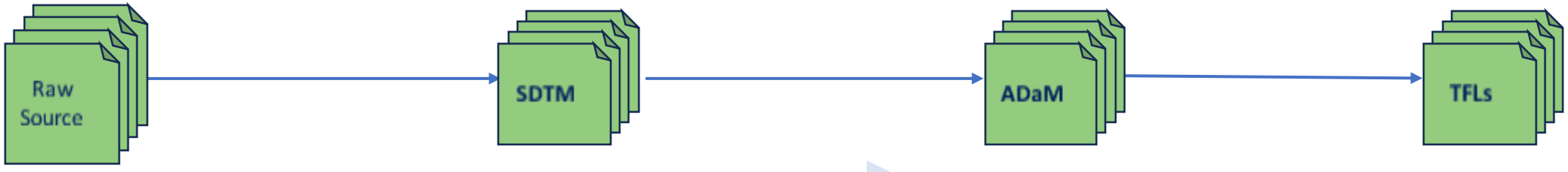
MDR Libraries for Standards, Terminologies and Mappings



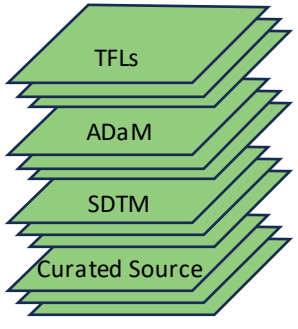
*
SPoT – Single Point of Truth
UDM - Unified, Longitudinally Integrated Subject Level Data Model
DSL – Data Services Layer (Analysis Ready Data post-cohort Selection)
TFLs – Tables, Figures, Listings

From Workflows Generating Files to Stacked Data in a Unified Repository

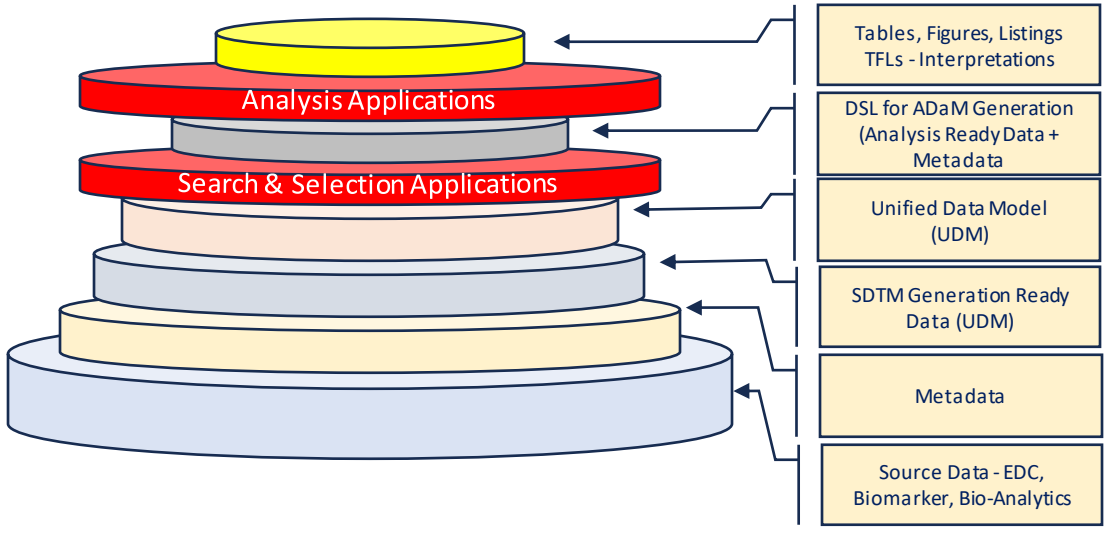
Conventional File-based Clinical Data Management and Analysis



Why not Stack files of Data managed by Biometrics?



Into a Multi-Layer Stack of Data, Metadata and Interpretations



Sample SAP and TFL Shell

2.3 Statistical Analysis Populations

The following sets will be used for the analyses performed in the study:

- *Enrolled Analysis Set (ENR):* All participants who have signed informed consent and enrolled into the study.
- *Randomized Analysis Set:* All participants who signed ICF and are assigned a randomization number.
- *Safety Analysis Set (SAF):* All participants who received the IV infusion (including partial infusion) of Drug or placebo will be included in the SAF.
- *PK Population (PK):* All participants who received Drug.
- *PK Analysis Population (PKAS):* All participants in the PK population for whom enough samples are available to determine at least one evaluable PK parameter after the single and/or multiple IV infusion of Drug and who have no major protocol deviations affecting the PK sampling and measurement will be included in PKAS.

ic Information and Baseline Characteristics, Safety Population (SAD)

			Cohort 3, 0.8 mg/kg N=# n(%)	Cohort 4, 1.2 mg/kg N=# n(%)	Cohort 7, 3 mg/kg N=# n(%)	Cohort 8, 6 mg/kg N=# n(%)	Cohort 9, 6 mg/kg (elderly) N=# n(%)
			#(##)	#(##)	#(##)	#(##)	#(##)
			#(##)	#(##)	#(##)	#(##)	#(##)
			#(##)	#(##)	#(##)	#(##)	#(##)
			##.# (#.#)	##.# (#.#)	##.# (#.#)	##.# (#.#)	##.# (#.#)
			## (##,##)	## (##,##)	## (##,##)	## (##,##)	## (##,##)
			# (###)	# (###)	# (###)	# (###)	# (###)
			# (##)	# (##)	# (##)	# (##)	# (##)
	American						
	White	# (##)	# (##)	# (##)	# (##)	# (##)	# (##)
-Ethnic		# (###)	# (###)	# (###)	# (###)	# (###)	# (###)
	Hispanic Or Latino	# (##)		# (##)	# (##)	# (##)	# (##)
	Not Hispanic Or Latino	# (##)	# (###)	# (##)	# (##)	# (##)	# (##)
-Height (cm)							
	Mean (SD)	###.## (##.##)	###.## (#.##)	###.# #(#.##)	###.# #(#.##)	###.# #(#.##)	###.# #(#.##)	###.# #(#.##)
	Median (min, max)	### (###,###)	### (###,###)	### (###,###)	### (###,###)	### (###,###)	### (###,###)	### (###,###)
-Weight (kg)							
	Mean (SD)	##.# (#.#)	##.# (##.#)	##.# (#.#)	##.# (#.#)	##.# (#.#)	##.# (#.#)	##.# (#.#)
	Median (min, max)	##.# (##.#,##.#)	##.## (##. #, ##.#)	##.## (##.#,##.#)	##.## (##.#,##.#)	##.## (##.#,##.#)	##.## (##.#,##.#)	##.## (##.#,##.#)

Abbreviations: N, number of subjects in treatment arm; n, number of subjects with given characteristic; SD, standard deviation

Program Name: XXXXXXXX Run Date: DDMMMYYYY HH:MM Clinical database last extracted: DDMMYYYY HH:MM

ADaM Derivation Examples

Domain	Variable	Algorithm Name	Algorithm Description
ADSL	AGEGR1	Alg.AGEGR1	Grouping of AGE into <50, 50-67, and >70
ADSL	AGEGR2	Alg.AGEGR2	Grouping of AGE into <50, >=50
ADSL	SAFFL	Alg.SAFFL	Y if ITTFL='Y' and TRTSDT ne missing. N otherwise
ADSL	ITTFL	Alg.ITTFL	Y if ARMCD ne ' '. N otherwise
ADSL	HEIGHTBL	Alg.HEIGHTBL	VS.VSSTRESN where VS.VSTESTCD='HEIGHT' and VS.VISITNUM=1
ADSL	TRTPREDT	Alg.PRETRTDATE	latest CM.CMENDTC where CM.CMCAT = 'ANTINEOPLASTIC THERAPY' or latest PR.PRENDTC where PR.PRCAT = 'RADIATION THERAPY'

ADaM Derivation - Formal Expressions (Executable)

Algorithm Name	Algorithm Description	Formal Expression (UDM)	Cohor
Alg.AGEGR1	Grouping of AGE into <50, 50-67, and >70	1: < 50 : {Subjects (AGE lt 50)} 2: >=50 and <=70 :{Subjects (AGE ge 50 and AGE le 70)} 3: > 70 : {Subjects (AGE gt 70)}	Yes
Alg.AGEGR2	Grouping of AGE into <50, >=50	1: < 50 : {Subjects (AGE lt 50)} 2: > 70 : {Subjects (AGE ge 50)}	Yes
Alg.SAFFL	Y if ITTFL='Y' and TRTSDT ne missing. N otherwise	Y : {Subjects (ARMCD ne ' ' and Subjects.RFXSTDTC ne '')}	Yes
Alg.ITTFL	Y if ARMCD ne ' '. N otherwise	Y : {Subjects (ARMCD ne ' ') }	Yes
Alg.HEIGHTBL	VS.VSSTRESN where VS.VSTESTCD='HEIGHT' and VS.VISITNUM=1	STRESN : {Findings (DOMAIN eq 'VS' AND TESTCD eq 'HEIGHT' and BLFL eq 'Y')}	Yes
Alg.PRETRTDATE	latest CM.CMENDTC where CM.CMCAT = 'ANTINEOPLASTIC THERAPY' or latest PR.PRENDTC where PR.PRCAT = 'RADIATION THERAPY'	{ Max(ENDTC) : Intervention (Domain eq 'CM' and CAT eq 'ANTINEOPLASTIC THERAPY') or (Domain eq 'PR' and CAT eq 'RADIATION THERAPY') }	Yes

ADaM Derivations from Cohorts and Search Expressions

Algorithm Name	Formal Expression (UDM)	Cohort?
Alg.AGEGR1	1: < 50 : {Subjects (AGE lt 50)} 2: >=50 and <=70 :{Subjects (AGE ge 50 and AGE le 70)} 3: > 70 : {Subjects (AGE gt 70)}	Yes
Alg.AGEGR2	1: < 50 : {Subjects (AGE lt 50)} 2: > 70 : {Subjects (AGE ge 50)}	Yes
Alg.SAFFL	Y : {Subjects (ARMCD ne ' ' and Subjects.RFXSTDTC ne '')}	Yes
Alg.ITTFL	Y : {Subjects (ARMCD ne ' ') }	Yes
Alg.HEIGHTBL	STRESN : {Findings (DOMAIN eq 'VS' AND TESTCD eq 'HEIGHT' and BLFL eq 'Y')}	Yes
Alg.PRETRTDATE	{ Max(ENDTC) : Intervention (Domain eq 'CM' and CAT eq 'ANTINEOPLASTIC THERAPY') or (Domain eq 'PR' and CAT eq 'RADIATION THERAPY') }	Yes

Num. Code	Value	Expression
1	"< 50"	{Subjects (AGE lt 50)}
2	">=50 and <=70"	{Subjects (AGE ge 50 and AGE le 70)}
3	"> 70"	{Subjects (AGE gt 70)}

Num. Code	Value	Expression
	STRESN	{ Findings (DOMAIN eq 'VS' AND TESTCD eq 'HEIGHT' and BLFL eq 1) }

Num. Code	Value	Expression
	Max(ENDTC)	{Intervention (Domain eq 'CM' and CAT eq 'ANTINEOPLASTIC THERAPY') or (Domain eq 'PR' and CAT eq 'RADIATION THERAPY') }

CDISC Analysis Results Metadata (ARM-TS) v1.0 for Define-XML v2.0

POINT CROSS Life Sciences

Documentation

references to external documents like the Statistical Analysis Plan (SAP), or the Protocol

Data References / Selection Criteria:

Parameters : <Test/Cat/…>
Selection criteria <visits/…>
Cohorts / Flags : <TRT*A/*FL/*GRy>

AnalysisReason :

CDISC Submission Value: ANLREAS, Code: C117744
"SPECIFIED IN PROTOCOL",
"SPECIFIED IN SAP",
"DATA DRIVEN",
"REQUESTED BY REGULATORY AGENCY"

AnalysisPurpose:

CDISC Submission Value: ANLPURP, Code: C117745:
"PRIMARY OUTCOME MEASURE",
"SECONDARY OUTCOME MEASURE",
"EXPLORATORY OUTCOME MEASURE"

Analysis Results Metadata (Detail) for Study CDISC-Sample

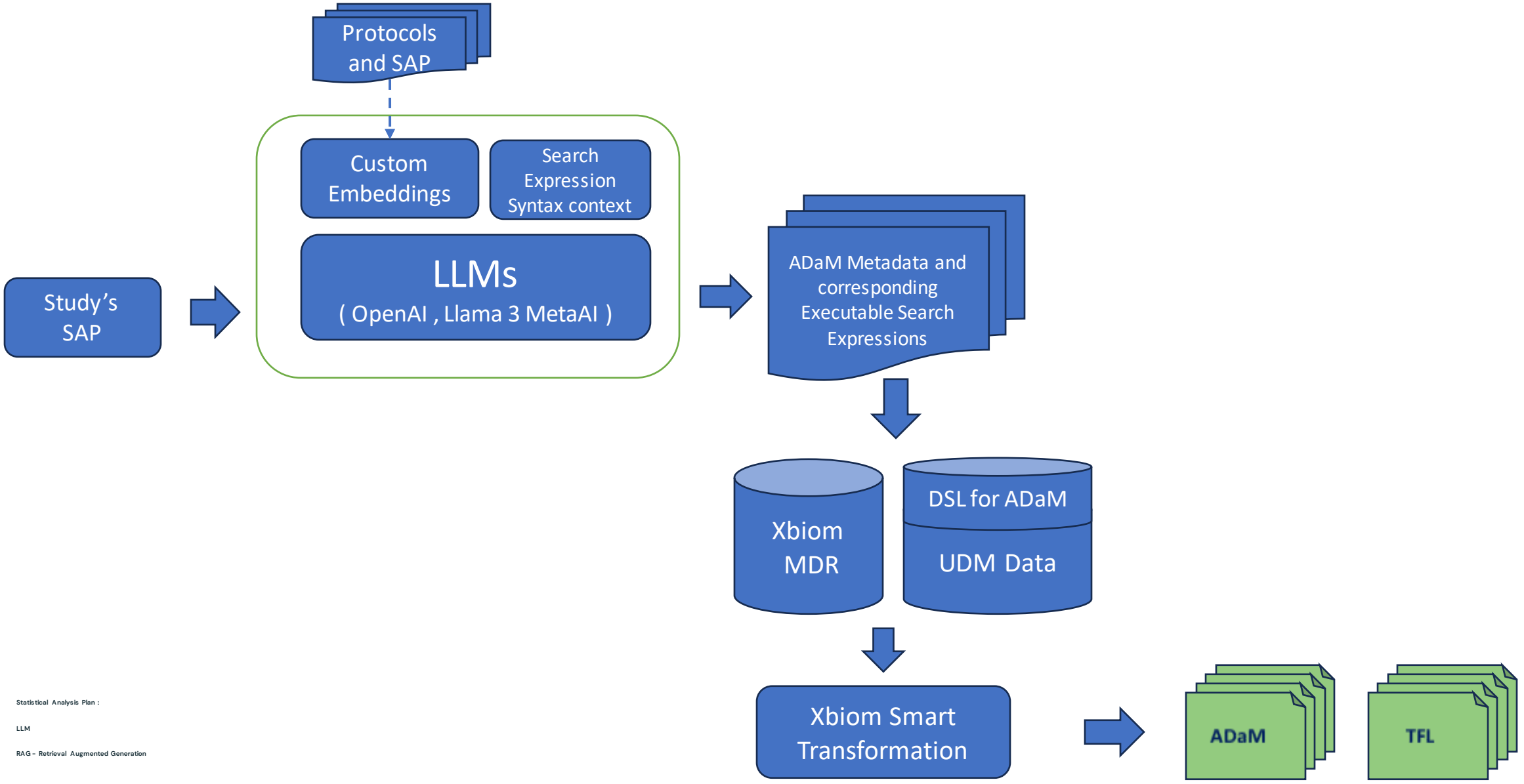
Table 14-3.01

Display	Table 14-3.01 **Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Populatio**
Analysis Result	**Dose response analysis for ADAS-Cog changes from baseline** (3)
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore) (4)
Analysis Variable(s)	CHG (Change from Baseline) (5)
Analysis Reason	SPECIFIED IN SAP (6)
Analysis Purpose	PRIMARY OUTCOME MEASURE (7)
Data References (incl. Selection Criteria)	ADQSADAS [PARAMCD = "ACTOT" and AVISIT = "Week 24" and EFFFL = "Y" and ANL01FL = "Y"] (9) (8)
Documentation	Linear model analysis of CHG for dose response; using randomized dose (0 for placebo; 54 for low dose; 81 for high dose) and site group in model. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose). (10) SAP Section 10.1.1 (11)
Programming Statements	[SAS version 9.2] (12)

```
proc glm data = ADQSADAS;
  where EFFFL='Y' and ANL01FL='Y' and AVISIT='Week 24' and PARAMCD="ACTOT";
  class SITEGR1;
  model CHG = TRTPN SITEGR1;
run;
```

https://www.cdisc.org/standards/foundational/define-xml/analysis-results-metadata-arm-v1-0-define-xml-v2-0

LLM Process for Creating Metadata for ADaM and TFL Generation



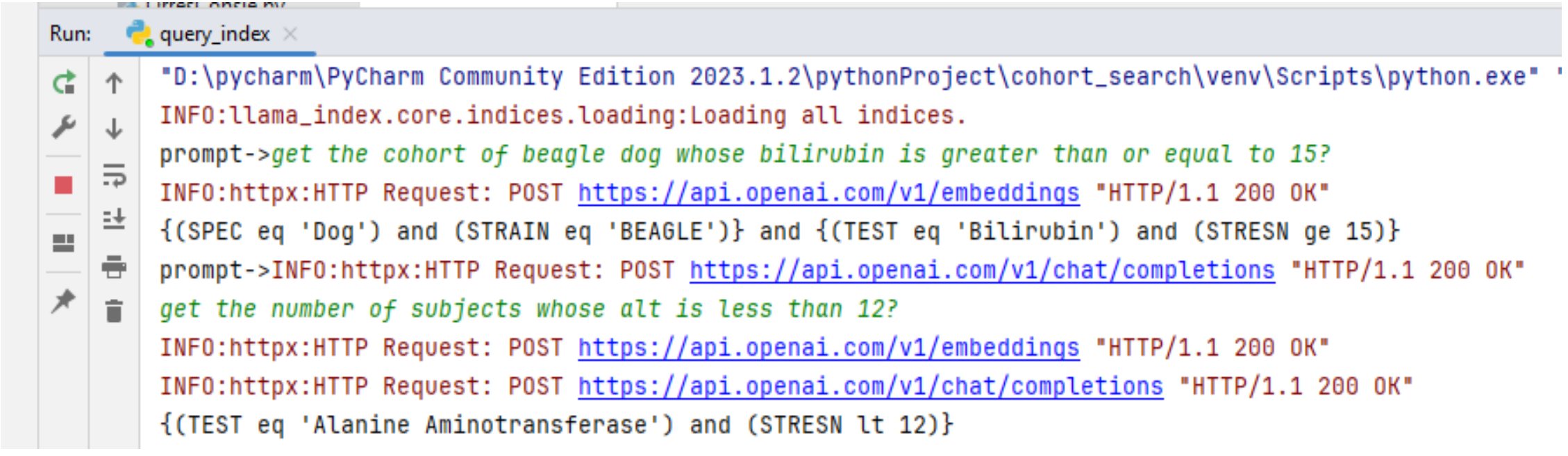
Statistical Analysis Plan :

LLM

RAG – Retrieval Augmented Generation

OpenAI

Syntax training



Thank you

Xbiom™ makes data useful

Quality is never an accident. It's always the result of intelligent effort
- John Ruskin